Estatística Inferencial
Intervalos de confiança e Testes de hipóteses

Distribuições multivariadas

1. Considere um vetor bivariado de variáveis aleatórias normalmente distribuído, com $\mathrm{E}(Y_1) = 0$, $\mathrm{E}(Y_2) = 4$, $\mathrm{V}(Y_1) = 1$, $\mathrm{V}(Y_2) = 9$ e $Cov(Y_1, Y_2) = 2$.
   a) Escreva a função de densidade probabilidade.
   b) Implemente tal densidade computacionalmente.
   c) Desenhe o gráfico da função de densidade bivariada.

2. Medidas de colesterol foram tomadas em um grande conjunto de pacientes que tiveram ataque do coração. Para cada paciente, medidas foram tomadas no dia 0, 2 e 4 dias após o ataque. Denote as respectivas v.a por X_0, X_2 e X_4. O vetor de médias foi $\boldsymbol{\mu_0} = (259.5, \mu_2 = 230.8, \mu_4 = 221.5)^\top$. A matriz de covariância é a seguinte:
$$\begin{bmatrix} 2276 & 1508 & 813 \\ 1508 & 2206 & 1349 \\ 813 & 1349 & 1865 \end{bmatrix}.$$
   a) Suponha que estamos interessados na distribuição de $X_0 - X_2$. Encontre esta distribuição sua média e variância. (Resposta: média 28.7 e variância 1466).
   b) Suponha que um paciente apresentou $\mu_0 = 260$. Calcule o valor esperado para X_2 e X_4 e suas respectivas variâncias. Forneça um intervalo de confiança com 95% de confiança.
   c) Calcule a probabilidade de X_4 ser maior que 270 para um paciente que chegou com colesterol de 250.
   d) Obtenha a matriz de correlação.

3. Um dado é lançado 12 vezes. Seja X_i o número de jogadas em que cada i caiu para cimaa, para $i = 1, \ldots, 6$.
   a) Calcule a esperança de X_i.
   b) Calcule a variância de X_i.
   c) Calcule a probabilidade de cada uma das faces cair para cima exatamente duas vezes.
   d) Implemente um código computacional ilustrando essa situação. Tente de forma aproximada calcular a probabilidade do item c).

Desigualdades

1. Suponha que a nota média dos alunos de Estatística Inferencial é de 70%. Dê um limite superior para a proporção de estudante que vão tirar nota de pelo menos 90%. Resposta: $\frac{7}{9}$.

2. Uma moeda é viciada de tal forma que a probabilidade de cair cara é de 0.20. Suponha que a moeda é lançada 20 vezes. Encotre um limite para a probabilidade de se obter ao mesmos 16 caras. Compare esse limite com a verdadeira probabilidade calculada usando a distribuição Binomial. Como você considera tal aproximação? Resposta: Limite $\frac{1}{4}$.

3. Considere uma variável aleatória X que toma o valor 0 com probabilidade $\frac{24}{25}$ e o valor 1 com probabilidade $\frac{1}{25}$. Calcule a $\mathrm{E}(X)$ e um limite superior para $\mathrm{P}(X \geq 5)$. Resposta: $\mathrm{P}(X \geq 5) = \frac{1}{25}$.
4. Suponha que uma moeda é lançada 100 vezes. Encontre um limite superior para a probabilidade do número de caras seja de no minimo 60 ou no máximo 40. Resposta: $\mathrm{E}(X) = 50$ e $\mathrm{V}(X) = 25$. $P(X < 40 \cup X > 60) = P(|X - \mu| \geq 10) \leq \frac{25}{10^2}$.

Lei dos grandes números e Teorema Central do limite

1. Seja $Y_1, \ldots, Y_n$ uma v.a iid da distribuição de Poisson com parâmetro λ.
   a) Mostre que a média amostral converge em probabilidade para λ quando $n \to \infty$.
   b) Encontre a distribuição aproximada da média amostral nesta situação.
   c) Faça uma ilustração computacional e compare a distribuição empírica com a distribuição aproximada.
2. Seja $Y_1, \ldots, Y_n$ v.a iid com $\mathrm{E}(Y_i) = \mu$ e $\mathrm{V}(Y_i) = \sigma^2 < \infty$.
   a) Mostre que $\overline{\sigma}^2 \xrightarrow{P} \sigma^2$ onde $\overline{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu)^2$.
   b) Obtenha a distribuição aproximada de $\overline{\sigma}^2$.
   c) Faça uma ilustração computacional da distribuição aproximada conforme o tamanho da amostra cresce. Use $n = 50$, 250 e 1000.
   d) Compare computacionalmente a distribuição aproximada com a distribuição exata apresentada na semana 3.
3. Sejam $Y_1, \ldots, Y_n$ v.a's iid com $\mathrm{E}(Y_i) = \mu$ e $\mathrm{V}(Y_i) = \sigma^2 < \infty$.
   a) Argumente sobre a seguinte afirmação: Para $n \to \infty$
   $$t = \frac{\overline{Y} - \mu}{s/\sqrt{n}} \xrightarrow{D} Z \sim N(0, 1).$$
   b) Faça uma ilustração computacional da distribuição aproximada conforme o tamanho da amostra cresce. Use $n = 50$, 250 e 1000.
   c) Compare computacionalmente a distribuição aproximada com a distribuição exata da estatística t apresentada na semana 3.
4. Sejam X_n e Y_n v.a's independentes com distribuição de Poisson de parâmetros n e m, respectivamente.
   a) Mostre que $R = \frac{(X_n - n) - (Y_n - m)}{\sqrt{X_n + Y_m}} \to Z \sim N(0, 1)$.
   b) Ilustre o resultado de a) computacionalmente e compare com a distribuição empírica de R.
   c) Comente sob os potenciais usos da estatística R na construção de teste de hipóteses, conforme discutido na semana 4.

Componentes dos modelos probabilísticos

1. Para cada uma das distribuições de probabilidade abaixo escreva a função de probabilidade ou densidade probabilidade, identifique o suporte, a esperança, a variância, os parâmetros e o espaço paramétrico.
   a) Distribuição Poisson de parâmetro λ.
   b) Distribuição binomial de parâmetros n e p.
   c) Distribuição exponencial de parâmetro λ.
   d) Distribuição normal de parâmetros μ e σ^2.
   e) Distribuição gama de parâmetros α e β.
   f) Distribuição uniforme de parâmetros a e b.
   g) Distribuição binomial negativa de parâmetros μ e ϕ.
   h) Distribuição log-normal de parâmetros μ e σ^2.
   i) Distribuição inversa Gaussiana de parâmetros μ e σ^2.
   j) Distribuição Tweeedie de parâmetros μ, ϕ e p.

Especificação de modelos

2. Para cada uma das situações abaixo proponha uma distribuição de probabilidade adequada e justifique sua escolha baseado em aspectos do fenômeno aleatório e características da distribuição. Descreva quais

aspectos da inferência estatística podem estar associados com cada uma das situações mencionadas.

a) Itens em uma linha de produção são classificados quanto a sua adequação aos padrões de produção. Apenas as condições conforme ou não-conforme são possíveis.
b) Uma pesquisa de mercado visa identificar o potencial de um novo negócio em uma cidade. Para isto um questionário com perguntas em uma escala likert de cinco níveis foi construído e aplicado a uma amostra de tamanho n da população de interesse.
c) Número de carros que chegam a um caixa automático de um banco durante um período de uma hora nas manhãs de fins de semana.
d) Ocorrência de defeitos relevantes em uma rodovia um mês após sua construção.
e) Medidas antropométricas (peso e altura) são tomadas em crianças do nono ano de escolas públicas brasileiras. Deseja-se caracterizar tais medidas para auxiliar na construção de equipamentos escolares de tamanho adequado.
f) Deseja-se estudar a distribuição do número de horas que um equipamento eletrônico funciona antes de apresentar defeitos com o objetivo de estabelecer um prazo razoável de garantia.
g) Número de quilômetros rodados que um novo pneu é capaz de rodar antes de apresentar defeitos.

Propriedades de estimadores

3. Seja $Y_i \sim B(n, p)$ para $i = 1, \ldots, n$ iid. Considere o estimador $\hat{p} = \frac{1}{n} \sum_{i=1}^{n} Y_i$ para p.

   a) Mostre que $\hat{p}$ é não-viciado para p.
   b) Encontre a variância de $\hat{p}$.
   c) Encontre o erro quadrático médio de $\hat{p}$.
   d) Mostre que $\hat{p}$ é médio quadrático consistente.
   e) Mostre que $\hat{p}$ é consistente em probabilidade.

4. Seja $Y_i \sim N(\mu, \sigma^2)$ para $i = 1, \ldots, n$ iid. Considere o estimador $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} Y_i$ para μ.

   a) Mostre que $\hat{\mu}$ é não-viciado para μ.
   b) Encontre a variância de $\hat{\mu}$.
   c) Encontre o erro quadrático médio de $\hat{\mu}$.
   d) Mostre que $\hat{\mu}$ é médio quadrático consistente.
   e) Mostre que $\hat{\mu}$ é consistente em probabilidade.

5. Seja $Y_i \sim N(\mu, \sigma^2)$ para $i = 1, \ldots, n$ iid. Considere o estimador $\hat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \mu)^2$ para μ. Note que μ é fixo e assumido como conhecido.

   a) $\hat{\sigma}^2$ é viciado para σ^2?
   b) Encontre a variância de $\hat{\sigma}$.
   c) Encontre o erro quadrático médio de $\hat{\sigma}^2$.
   d) Mostre que $\hat{\sigma}^2$ é médio quadrático consistente.
   e) Mostre que $\hat{\sigma}^2$ é consistente em probabilidade.

6. Seja $Y_i \sim N(\mu, \sigma^2)$ para $i = 1, \ldots, n$ iid. Considere o estimador $\hat{\sigma^2} = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \hat{\mu})^2$ para σ^2.

   a) Mostre que $\hat{\sigma}^2$ é não-viciado para σ^2.
   b) Encontre a variância de $\hat{\sigma}^2$.
   c) Encontre o erro quadrático médio de $\hat{\sigma}^2$.
   d) Mostre que $\hat{\sigma}^2$ é médio quadrático consistente.
   e) Mostre que $\hat{\sigma}^2$ é consistente em probabilidade.

7. Sejam $Y_i \sim N(\mu, \sigma^2)$, para $i = 1, \ldots, n$.

   a) Mostre que o estimador $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} Y_i$ é não viciado para μ.
   b) Mostre que o estimador $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\mu})^2$ é viciado para σ^2 e determine o seu viés.
   c) Mostre que o estimador $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \hat{\mu})^2$ é não viciado para σ^2.
   d) Compare os estimadores para σ^2 em b) e c) em termos assintóticos. O que você pode concluir?

8. Mostre que a média das duas primeiras observações de um conjunto de n observações independentes é não-viciado mas não é consistente, para estimar a média populacional.

9. Determine a condição dos coeficientes a_i, de tal forma que a combinação linear $\sum a_i Y_i$ seja não viciada para $\mathrm{E}(Y)$.

10. Mostre que se T é uma estimativa não viciada para θ, então $aT + b$ é uma estimativa não-viciada de $a\theta + b$. E T^2 é uma estimativa não-viciada para θ^2? Justifique.

11. Seja Y o número de sucessos em n ensaios Bernoulli com parâmetro p. Determine o erro quadrático médio na estimação de p por cada um dos estimadores $T_1 = Y/n$ e $T_2 = (Y+1)/(n+2)$. Um estimador é melhor do que outro?

Verossimilhança e estimação pontual

12. Para cada uma das distribuições de probabilidade abaixo escreva a função de verossimilhança e log-verossimilhança.
    a) Distribuição Poisson de parâmetro λ.
    b) Distribuição binomial de parâmetros n e p.
    c) Distribuição exponencial de parâmetro λ.
    d) Distribuição normal de parâmetros μ e σ^2.
    e) Distribuição gama de parâmetros α e β.
    f) Distribuição uniforme de parâmetros a e b.
    g) Distribuição binomial negativa de parâmetros μ e ϕ.
    h) Distribuição log-normal de parâmetros μ e σ^2.
    i) Distribuição inversa Gaussiana de parâmetros μ e σ^2.
    j) Distribuição Tweeedie de parâmetros μ, ϕ e p.
13. Para cada uma das situações abaixo encontre o estimador de máxima verossimilhança.
    a) Distribuição Poisson de parâmetro λ.
    b) Distribuição binomial de parâmetros n(conhecido) e p.
    c) Distribuição exponencial de parâmetro λ.
    d) Distribuição normal de parâmetros μ e σ^2(conhecido).
    e) Distribuição normal de parâmetros μ(conhecido) e σ^2.
    f) Distribuição gama de parâmetros α(conhecido) e β.
    g) Distribuição gama de parâmetros α e β(conhecido).
    h) Distribuição gama de parâmetros α e β(conhecido).
    i) Distribuição binomial negativa de parâmetros μ e ϕ(conhecido).
    j) Distribuição log-normal de parâmetros μ e σ^2.

Verossimilhança e suas derivadas

1. Para cada uma das distribuições de probabilidade abaixo (quando possível) escreva a função de verossimilhança, log-verossimilhança, escore, informação observada e informação esperada. Para cada distribuição quando possível encontre o limite inferior da variância de Cramér-Rao.
    a) Distribuição Poisson de parâmetro λ.
    b) Distribuição binomial de parâmetros n e p.
    c) Distribuição exponencial de parâmetro λ.
    d) Distribuição normal de parâmetros μ e σ^2.
    e) Distribuição gama de parâmetros α e β.
    f) Distribuição uniforme de parâmetros a e b.
    g) Distribuição binomial negativa de parâmetros μ e ϕ.
    h) Distribuição log-normal de parâmetros μ e σ^2.
    i) Distribuição inversa Gaussiana de parâmetros μ e σ^2.
    j) Distribuição Tweeedie de parâmetros μ, ϕ e p. Considere que o p é conhecido.

Distribuição assintótica do EMV

1. Para cada uma das distribuições de probabilidade abaixo (quando possível) escreva a função de verossimilhança, log-verossimilhança, escore, informação observada e informação esperada. Para cada distribuição quando possível encontre o limite inferior da variância de Cramér-Rao.
   a) Distribuição Poisson de parâmetro λ.
   b) Distribuição binomial de parâmetros n e p.
   c) Distribuição exponencial de parâmetro λ.
   d) Distribuição normal de parâmetros μ e σ^2.
   e) Distribuição gama de parâmetros α e β.
   f) Distribuição uniforme de parâmetros a e b.
   g) Distribuição binomial negativa de parâmetros μ e ϕ.
   h) Distribuição log-normal de parâmetros μ e σ^2.
   i) Distribuição inversa Gaussiana de parâmetros μ e σ^2.
   j) Distribuição Tweeedie de parâmetros μ, ϕ e p. Considere que o p é conhecido.

Nos casos que não foi possível resolver analiticamente, resolva numéricamente.

Intervalos de confiança

1. Para cada uma das distribuições de probabilidade abaixo encontre o EMV e um intervalo de confiança usando pelo menos duas estratégias. Note que implementação computacional será necessário.
   a) Distribuição Poisson de parâmetro λ.
   b) Distribuição binomial de parâmetros n e p.
   c) Distribuição exponencial de parâmetro λ.
   d) Distribuição normal de parâmetros μ e σ^2.
   e) Distribuição gama de parâmetros α e β.
   f) Distribuição log-normal de parâmetros μ e σ^2.
   g) Distribuição inversa Gaussiana de parâmetros μ e σ^2.
   h) Distribuição Tweeedie de parâmetros μ, ϕ e p. Considere que o p é conhecido.

Testes de hipóteses

1. Para cada uma das distribuições de probabilidade construa a estatística de teste conforme solicitado.
   a) Distribuição Poisson de parâmetro λ. Construa um teste para avaliar se $\lambda = \lambda_0$ contra $\lambda \neq \lambda_0$. Faça um exemplo com dados simulados.
   b) Distribuição binomial de parâmetros n e p. Construa um teste para avaliar se $p = 0.5$ contra $p \neq 0.5$. Faça um exemplo com dados simulados. Para o valor de $p = 0.8$ calcule o poder do teste.
   c) Distribuição exponencial de parâmetro λ. Construa o teste da razão de verossimilhança para avaliar se $\lambda = \lambda_0$ contra $\lambda \neq \lambda_0$. Forneça uma implementação computacional.
   d) Distribuição normal de parâmetros μ e σ^2. Construa um teste escore para testar se $\sigma = 1$ contra $\sigma \neq 1$. Faça um estudo de simulação para verificar a qualidade do teste proposto. Você teria uma sugestão para melhorar este teste?
   e) Distribuição Tweeedie de parâmetros μ, ϕ e p. Considere que o p é conhecido. Construa um teste Wald para testar se $\phi = 1$ contra $\phi \neq 1$. Faça um estudo de simulação para verificar a qualidade do teste proposto. Você teria uma sugestão para melhorar este teste?